

Editorial

Validación del EuroSCORE II en España. ¿Y ahora qué hacemos?



Validation of the EuroSCORE II in Spain. Now what do we do?

Jacobo Silva Guisasaola

Servicio de Cirugía Cardíaca Hospital Universitario Central de Asturias, Oviedo, Asturias, España

Es innegable que las escalas de riesgo son una herramienta indispensable hoy en día para llevar a cabo la actividad asistencial en un servicio de cirugía cardíaca. Dicha herramienta puede ser utilizada para estimar el riesgo en un paciente concreto o como control de calidad a la hora de validar los resultados globales de un centro. El modelo EuroSCORE fue desde su inicio una escala que rápidamente se extendió en su uso entre otras razones por sus características parsimoniosas. También es innegable que el EuroSCORE fue diseñado en los años 90 y muchos grupos quirúrgicos detectaron desde hacía años una infraestimación de la mortalidad con esta escala, sobre todo en determinadas patologías¹. Así se hacía necesaria una actualización del mismo, hecho que se llevó a cabo con el proyecto EuroSCORE II (EII)².

Desde la publicación de sus resultados, llamó la atención la baja mortalidad predicha que asignaba esta nueva escala. Así, un paciente varón de 82 años diagnosticado de estenosis aórtica severa y en clase funcional II de la NYHA, sin otras comorbilidades asociadas, tiene un riesgo asignado de muerte según el EII del 1,04%. Se han apuntado algunas deficiencias metodológicas en el diseño del mismo, como la ausencia de variables de interacción, la falta de información sobre los casos perdidos, o que la mortalidad fue computada únicamente como la hospitalaria y no esta más la mortalidad a los 30 días. Dichas deficiencias cuestionan de manera muy importante su validez externa según algunos grupos^{3–18} (tabla 1). Al menos 12 estudios apuntan a una mala calibración global o en determinados subgrupos como los pacientes valvulares y/o de alto riesgo^{4–18}. Sin embargo, muchos de estos estudios han de ser interpretados con cautela, ya que tienen un diseño retrospectivo y/o incluyen un número reducido de pacientes. Creo que sería erróneo y precipitado descartar la validez de esta nueva escala desarrollada a partir de una población tan numerosa y diseñada por un grupo con una experiencia muy importante en la estimación del riesgo preoperatorio en cirugía cardíaca².

Antes de aplicar una escala de riesgo debemos conocer si esta funciona en nuestros pacientes, hecho que se conoce como validación externa. En este sentido, el artículo de García-Valentín et al.¹⁹ constituye una aportación muy valiosa en nuestro país para la aplicación futura de la escala EII. Las características prospectivas de este estudio, el hecho de ser multicéntrico y, además, poseer un porcentaje del 100% en la cumplimentación de los datos le dan una relevancia científica incuestionable. Además, la concordancia de los resultados con numerosos estudios^{4–18}, alguno de ellos desarrollado también en nuestro país⁷, le da aún más relevancia e interés si cabe. Creo justo felicitar a todos los centros participantes por la respuesta a la propuesta y por la calidad en la recogida de los

datos. Esta iniciativa, no exenta de dificultades, debe ser puesta como ejemplo para futuros proyectos en nuestra sociedad.

Este estudio multicéntrico, en el que intervinieron 20 centros de nuestro país, muestra una mortalidad observada del 6,5% y una mortalidad esperada según el EuroSCORE logístico del 9,8% y el 5,7% según el EII. Los hallazgos fundamentales son una buena discriminación del EII (área bajo la curva ROC 0,79 vs. EuroSCORE clásico 0,77; diferencias no significativas) y una mala calibración global¹⁹. El método empleado para el estudio de la calibración fue el de Hosmer-Lemeshow, que aplica un test de la chi al cuadrado a los diferentes deciles de la muestra, comparando la mortalidad observada con la predicha por el modelo. Los coeficientes chi al cuadrado fueron 33,02 para el EuroSCORE clásico y 38,98 para el EII. Se puede afirmar de forma grosera que el EII tiene incluso peor calibración en este estudio, ya que el valor de este coeficiente era mayor. Dicho test ha sufrido muchas críticas como herramienta de calibración, sobre todo en muestras grandes, ya que su sensibilidad puede ser muy elevada²⁰. Es claro que la calibración de un modelo no puede ser presentada de forma dicotómica, como algo blanco o negro y probablemente haya una escala de grises. Existen otras pruebas para valorar la calibración y otros estadísticos que aportan información muy valiosa. Estos no han sido incluidos clásicamente en los trabajos de validación de las escalas de riesgo en cirugía cardíaca. Así, la escala de Brier o las curvas de regresión para la calibración y el coeficiente R² que aporta información sobre el porcentaje de mortalidad que explica el modelo deberían ser incluidos en dichos trabajos. Por último, se ha desarrollado en los últimos años una nueva metodología para la validación de los modelos predictivos²¹. Herramientas como las tablas de reclasificación y los llamados «*net reclassification improvement (NRI)*» e «*integrated discrimination improvement (IDI)*» aportan un nuevo enfoque en la validación de los modelos predictivos; basándose en establecer varias categorías dentro de cada variable predictora y crear varios subgrupos de riesgo en sobre la base de esta nueva clasificación. La utilización de esta nueva metodología esclarecerá muchas de las dudas que hoy en día plantean las escalas de riesgo en cirugía cardíaca.

Se pueden intentar buscar muchas explicaciones a las discrepancias encontradas entre este trabajo y el estudio original del EII, pero, sin lugar a duda, la razón fundamental de estos hallazgos son las diferencias poblacionales y de patología de base existentes entre la muestra del EII y la población de nuestro país. En el estudio de García-Valentín et al. el porcentaje de pacientes con cirugía coronaria asilada fue del 25% frente al 46,7% en el registro del EII. La puntuación media del EII en este estudio fue de 5,7% frente al 3,9% en el registro de validación interna del proyecto EII. Varios trabajos publicados recientemente muestran una peor calibración del EII en pacientes valvulares y/o de riesgo elevado; precisamente, este suele ser el grupo mayoritario de pacientes que se operan

Correo electrónico: jsilva8252@yahoo.es

Tabla 1

Principales estudios publicados sobre la validación del EuroSCORE II

Autor	Período	n	Procedimiento	ROC	Calibración
Nashef et al. ²	2010	22.381	C. cardíaca	0,81	p=0,051
Di Dedda et al. ³	2011	1.090	C. cardíaca	0,81	p=0,22
Chalmers et al. ⁴	2006-2011	5.576	C. cardíaca	0,79	p<0,001
Grant et al. ⁵	2010-2011	23.740	C. cardíaca	0,808	p=0,003
Barili et al. ⁶	2006-2011	12.325	C. cardíaca	0,82	p<0,005
Carnero ⁷	2005-2010	3.798	C. cardíaca	0,85	p<0,001
Kirmani et al. ⁸	2001-2010	15.499	C. cardíaca	0,82	p<0,001
Borde et al. ⁹	2011-2012	498	CABG + valv.	0,69	p=0,71
Biancari et al. ¹⁰	2006-2011	1.027	CABG	0,85	
Kunt et al. ¹¹	2004-2012	428	CABG	0,72	p<0,05
Zhang et al. ¹²	2006-2011	3.479	Valvular	0,68	p<0,001
Wang et al. ¹³	2008-2011	11.170	Valvular	0,72	p=0,002
Howell et al. ¹⁴	2006-2011	933	Alto Riesgo	0,67	p=0,035
Paparella et al. ¹⁵	2011-2012	6.293	C. cardíaca.Subgrupo alto riesgo	0,83	Buena calibración.Infraestimación alto riesgo
Borracci et al. ¹⁶	2012-2013	503	C. cardíaca	0,856	p=0,082
Qadir et al. ¹⁷	2006-2010	2.004	CABG	0,835	p=0,013
Kosztá et al. ¹⁸		2.287	C. cardíaca	0,818	p=0,0084

en muchos servicios de nuestro país, con un porcentaje llamativamente pequeño de patología coronaria si nos comparamos con el resto de Europa²².

Las preguntas que surgen tras conocer los resultados de este estudio son claras: ¿y ahora qué hacemos?, ¿seguimos aplicando el EuroSCORE clásico? Creo que encontrar las respuestas a estas 2 preguntas no resulta tarea fácil, pero sin duda el sentido común nos debe llevar a pensar que lo correcto sería realizar una validación externa en cada servicio, si existe una correcta recogida de los datos, con un control de calidad y con un tamaño suficiente de la muestra que se podría estimar en más de 1.000 pacientes. Si los resultados muestran una buena calibración y capacidad de discriminación, se aplicaría directamente el EII. Si, por el contrario, existe una mala calibración, sobre todo en los subgrupos de alto riesgo y/o valvulares, se debería aplicar el EII para los grupos de bajo-mediano riesgo y continuar con el EuroSCORE clásico para los de alto riesgo.

Por último, me gustaría hacer una última reflexión sobre la aplicación de las escalas de riesgo en nuestro país sin pecar de autocomplaciente. Nos debemos liberar del complejo de malos resultados que hemos tenido en los últimos años, sobre todo cuando nos comparaban de forma cruda con otros países o cuando esa mortalidad era ajustada con escalas que no tenían una validación correcta en nuestros pacientes. Este argumento ha servido para cuestionar las indicaciones quirúrgicas que estaban establecidas de forma clara en las guías, sobre todo en cirugía coronaria. Sin embargo, los resultados de la cirugía cardíaca en España han experimentado una mejoría notable en los últimos años. Un ejemplo claro de que nuestros resultados no son tan malos, o incluso se podría decir que son muy buenos, son los datos que se desprenden del Primer Informe del Proyecto Español de Calidad de Cirugía Cardiovascular del Adulto 2013. Este proyecto mostró que la mortalidad ajustada por riesgo (mediante un modelo propio) en cirugía coronaria fue inferior a la de países como Alemania o Gran Bretaña²².

Bibliografía

- Parolari A, Pesce LL, Trezzi M, Cavallotti L, Kassem S, Loardi C, et al. EuroSCORE performance in valve surgery: A meta-analysis. *Ann Thorac Surg.* 2010;89(3):787-93.
- Nashef SA, Roques F, Sharples LD, Nilsson J, Smith C, Goldstone AR, et al. Euroscore II. *Eur J Cardiothorac Surg.* 2012;41(4):734-44.
- Di Dedda U, Pelissero G, Agnelli B, De Vincentiis C, Castelvichio S, Ranucci M. Accuracy, calibration and clinical performance of the new EuroSCORE II risk stratification system. *Eur J Cardiothorac Surg.* 2013;43(1):27-32.
- Chalmers J, Pullan M, Fabri B, McShane J, Shaw M, Mediratta N, et al. Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. *Eur J Cardiothorac Surg.* 2013;43(4):688-94.
- Grant SW, Hickey GL, Dimarakis I, Trivedi U, Bryan A, Treasure T, et al. Bridge-water BHow does EuroSCORE II perform in UK cardiac surgery; an analysis of 23 740 patients from the Society for Cardiothoracic Surgery in Great Britain and Ireland National Database. *Heart.* 2012;98(21):1568-72.
- Barili F, Pacini D, Capo A, Rasovic O, Grossi C, Alamanni F, et al. Does EuroSCORE II perform better than its original versions? A multicentre validation study. *Eur Heart J.* 2013;34(1).
- Carnero-Alcázar M, Silva Guisasola JA, Reguillo Lacruz FJ, Maroto Castellanos LC, Cobiella Carnicer J, Villagrán Medinilla E, et al. Validation of EuroSCORE II on a single-centre 3800 patient cohort. *Interact Cardiovasc Thorac Surg.* 2013;16(3):293-300.
- Kirmani BH, Mazhar K, Fabri BM, Pullan DM. Comparison of the EuroSCORE II and Society of Thoracic Surgeons 2008 risk tools. *Eur J Cardiothorac Surg.* 2013;44(6):999-1005.
- Borde D, Gandhe U, Hargave N, Pandey K, Khullar V. The application of European system for cardiac operative risk evaluation II (EuroSCORE II) and Society of Thoracic Surgeons (STS) risk-score for risk stratification in Indian patients undergoing cardiac surgery. *Ann Card Anaesth.* 2013;16(3):163-6.
- Biancari F, Vasques F, Mikkola R, Martin M, Lahtinen J, Heikkinen J. Validation of EuroSCORE II in patients undergoing coronary artery bypass surgery. *Ann Thorac Surg.* 2012;93(6):1930-5.
- Kunt AG, Kurtcephe M, Hidioglu M, Cetin L, Kucuker A, Bakuy V, et al. Comparison of original EuroSCORE, EuroSCORE II and STS risk models in a Turkish cardiac surgical cohort. *Interact Cardiovasc Thorac Surg.* 2013;16(5):625-9.
- Zhang GX, Wang C, Wang L, Lu FL, Li BL, Han L, et al. Validation of EuroSCORE II in Chinese patients undergoing heart valve surgery. *Heart Lung Circ.* 2013;22(8):606-11.
- Wang L, Han QQ, Qiao F, Wang C, Zhang XW, Han L, et al. Performance of EuroSCORE II in patients who have undergone heart valve surgery: A multicentre study in a Chinese population. *Eur J Cardiothorac Surg.* 2014;45(2):359-64.
- Howell NJ, Head SJ, Freemantle N, van der Meulen TA, Senanayake E, Menon A, et al. The new EuroSCORE II does not improve prediction of mortality in high-risk patients undergoing cardiac surgery: A collaborative analysis of two European centres. *Eur J Cardiothorac Surg.* 2013;44(6):1006-11.
- Paparella D, Guida P, di Eusanio G, Caparrotti S, Gregorini R, Cassese M, et al. Fiore TRisk stratification for in-hospital mortality after cardiac surgery: External validation of EuroSCORE II in a prospective regional registry. *Eur J Cardiothorac Surg.* 2014;46:840-8.
- Borracci RA, Rubio M, Celano L, Ingino CA, Allende NG, Ahuad Guerrero RA. Prospective validation of EuroSCORE II in patients undergoing cardiac surgery in Argentinean centres. *Interact Cardiovasc Thorac Surg.* 2014;18(5):539-43.
- Qadir I, Alamzaib SM, Ahmad M, Perveen S, Sharif H. EuroSCORE vs EuroSCORE II vs. Society of Thoracic Surgeons risk algorithm. *Asian Cardiovasc Thorac Ann.* 2014;22(2):165-71.
- Kosztá G, Sira G, Szatmári K, Farkas E, Szeráfin T, Fülesdi B. Performance of EuroSCORE II in Hungary: A single-centre validation study. *Heart Lung Circ.* 2014. <http://dx.doi.org/10.1016/j.hlc.2014.04.005>. S1443-9506(14)00178-4.
- García-Valentín A, Bernabeu E, Pereda D, Josa M, Cortina JM, Mestres CA, et al. Validación de EuroSCORE II en España. *Cir Cardiov.* 2014;21(4):246-51.
- Nezic D, Borzanovic M, Spasic T, Vukovic P. Calibration of the EuroSCORE II risk stratification model: Is the Hosmer-Lemeshow test acceptable any more? *Eur J Cardiothorac Surg.* 2013;43(1):206.
- Steyerberg EW, Vickers AJ, Cook NR, Gerdts T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology.* 2010;21(1):128-38.
- Primer Informe del Proyecto Español de Calidad de Cirugía Cardiovascular del Adulto 2013. United Kingdom: Dendrite Clinical Systems Ltd Oxfordshire.